

Amendments to the Claims

1. (currently amended) A computer-implemented method for improving information retrieval, classification, indexing, and summarization, comprising:
identifying ~~a compound document as a coherent body of a collection of~~ hyperlinked material documents as a single coherent compound document on a single topic as created by a number of collaborating authors, wherein the identifying includes observing results of a first number of heuristics run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates;
analyzing the content and structure of the compound document to find a preferred entry point for the compound document, wherein the analyzing includes observing results of a second number of heuristics run on the compound document and related hyperlinks, wherein the analyzing includes combining the results of the second number of heuristics run on various hyperlinked documents of the compound document, wherein the results of the second number of heuristics include numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point;
processing the compound document as a whole, including at least one of indexing, classification, and retrieval; and
processing the compound document from the entry point, including at least one of creating at least one of presentation of results from retrieval, summarization, and classification.
2. (currently amended) The method of claim 1 wherein the ~~body~~ collection of hyperlinked ~~material documents~~ includes material from at least one of: the internet, an intranet, and a digital library.
3. (currently amended) The method of claim 1 wherein the ~~body~~ collection of hyperlinked ~~material documents~~ is distributed over a plurality of URLs.

4. (canceled)

5. (currently amended) The method of ~~claim 4~~claim 1 wherein the first number of heuristics includes identifying hyperlinks that link within the same directory and include a sufficient quantity of common anchor text.

6. (currently amended) The method of ~~claim 4~~claim 1 wherein the first number of heuristics includes identifying hyperlinks that contain linguistic structures that indicate relationships between parts of a document including at least one of a list of page numbers, and the terms “next”, “previous”, “index”, “contents”, and their non-English equivalents.

7. (currently amended) The method of ~~claim 4~~claim 1 wherein the first number of heuristics includes identifying external hyperlinks to the same places.

8. (canceled)

9. (currently amended) The method of ~~claim 4~~claim 1 wherein the first number of heuristics includes identifying individual URLs having similar structure indicating an order of inclusion in the compound document.

10. (currently amended) The method of ~~claim 4~~claim 1 wherein the first number of heuristics includes identifying a link structure of “wheel” form.

11. (canceled)

12. (currently amended) The method of ~~claim 11~~claim 1 wherein the second number of heuristics includes identifying specific filenames that define the entry point, including at least one of: “index” and “default”.

13. (currently amended) The method of ~~claim 11~~ claim 1 wherein the second number of heuristics includes identifying a particular component document in the compound document as the entry point because the component document has several in-links.

14. (original) The method of claim 13 wherein the in-links are from outside the compound document.

15. (currently amended) The method of ~~claim 11~~ claim 1 wherein the second number of heuristics includes identifying a particular component document in the compound document as the entry point because the component document has several out-links.

16. (currently amended) The method of ~~claim 11~~ claim 1 wherein the second number of heuristics includes determining a measure of vector distances along intra-document links between a particular component document and all other component documents in the compound document.

17. (currently amended) The method of ~~claim 11~~ claim 1 wherein the second number of heuristics includes determining whether a URL has links pointing to longer URLs having common directory components followed by different ending directory components.

18. (original) The method of claim 17 wherein the ending directory components contain specific identifying information.

19. (canceled)

20. (canceled)

21. (canceled)